Extreme values of particular nonlinear processes

Dominique Guegan, Sophie A. Ladoucette

Extreme values of particular non linear processes

Dominique GUÉGAN [a], Sophie LADOUCETTE [a,b]

[a] Ecole Normale Supérieure de Cachan, GRID, UMR CNRS 8534, 61 avenue du Président Wilson, 94235 Cachan cedex, France.

[b] Banque de France, 41-1391 DGEI-DEER-Centre de Recherche, 31 rue Croix des Petits Champs, 75049 Paris cedex 1, France.
E-mail : guegan@grid.ens-cachan.fr, sophie.ladoucette@banque-france.fr

Abstract. We are interested in the behavior of the maxima of a general class of deterministic chaotic processes - including the tent map and the logistic map -, of noisy chaotic processes, and of the long memory Gaussian k-factor Gegenbauer processes.

Valeurs extrêmes pour des processus non linéaires particuliers

Résumé. *Nous nous intéressons au comportement des maxima d'une classe générale de processus chaotiques déterministes - comprenant les applications tent et logistique -, de processus chaotiques bruités et des processus à mémoire longue de Gegenbauer à k-facteurs Gaussiens.*

Version française abrégée

Dans cette Note, nous nous intéressons au comportement asymptotique des maxima de certains processus dépendants. Nous considérons d'abord les processus générés par une famille paramétrique d'applications chaotiques définie par :

$$X_{n+1} = \varphi_\nu(X_n) = \begin{cases} 1-(1-2X_n)^\nu, \ 0 \le X_n < 1/2 \\ 1-(2X_n-1)^\nu, \ 1/2 \le X_n \le 1 \end{cases} \tag{3}$$

où $1/2 \le \nu \le 2$ (voir Hall et Wolff [9] et Lawrance et Spencer [10]). Nous considérons aussi les processus générés par une version bruité du système chaotique (3) pris en $\nu = 2$, définie par :

$$\begin{cases} Y_n = X_n + \varepsilon_n, \ n \in \mathbb{N} \\ X_n = 4X_{n-1}(1-X_{n-1}), \ n \in \mathbb{N}^* \end{cases} \tag{4}$$

avec $(\varepsilon_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées de densité $\beta(a,b)$ $(a,b>0)$, indépendantes de $(X_n)_{n\in\mathbb{N}}$. Enfin, nous nous intéressons aux processus de Gegenbauer à k-facteurs Gaussiens définis par :

$$\prod_{i=1}^{k}(I-2\upsilon_i B+B^2)^{d_i}X_n = \varepsilon_n \tag{5}$$

avec $|d_i|<1/2$ si $|\upsilon_i|<1$, $|d_i|<1/4$ si $|\upsilon_i|=1$, $d_i \neq 0$, pour $i=1,\cdots,k$, et $(\varepsilon_n)_{n\in\mathbb{Z}}$ un bruit blanc Gaussien (voir Gray *et al.* [6] et Giraitis et Leipus [4]).

En établissant précisément la distribution limite de leurs maxima, nous généralisons le Théorème 1 de Fisher-Tippett [3] pour les processus dépendants (3), (4) et (5).

Dans la suite, nous noterons $(M_n)_{n\in\mathbb{N}}$ la suite des maxima $M_n = \max(Z_1, \cdots, Z_n)$, $n \geq 1$, associée à un processus $(Z_n)_{n\in\mathbb{N}}$ de distribution marginale F.

Dans les Théorèmes 2 et 3, nous donnons les domaines d'attraction des distributions marginales associées aux processus (3) et (4), sous l'hypothèse de non nullité de l'index extrémal θ défini en (2).

THÉORÈME 2. – *Soit $(X_n)_{n\in\mathbb{N}}$ un processus associé au système (3) avec ν fixé, dont la fonction de densité f est supposée continue et strictement positive en $1/2$. Si $(X_n)_{n\in\mathbb{N}}$ possède un index extrémal vérifiant $0 < \theta \leq 1$, alors F appartient au domaine d'attraction de Weibull d'index $1/\nu$, i.e. :*

$$\mathbb{P}\{c_n^{-1}(M_n - d_n) \leq x\} \to \Psi_{1/\nu}(x)$$

avec $c_n = (f(1/2)\theta n)^{-\nu}$ et $d_n = 1$.

THÉORÈME 3. – *Soit $(Y_n)_{n\in\mathbb{N}}$ un processus associé au système (4) avec $(\varepsilon_n)_{n\in\mathbb{N}}$ de densité $\beta(a,b)$. Si $(Y_n)_{n\in\mathbb{N}}$ possède un index extrémal vérifiant $0 < \theta \leq 1$, alors F appartient au domaine d'attraction de Weibull d'index $(2b+1)/2$, i.e. :*

$$\mathbb{P}\{c_n^{-1}(M_n - d_n) \leq x\} \to \Psi_{(2b+1)/2}(x)$$

avec $c_n = (\frac{4\Gamma(a+b)\theta n}{(2b+1)\pi\Gamma(a)\Gamma(b)})^{-2/(2b+1)}$ et $d_n = 2$, où Γ est la fonction Gamma.

Nous établissons maintenant un résultat similaire pour le processus (5), en remarquant que la fonction d'autocovariance γ de ce processus vérifie la condition de Berman [1] $\gamma(n)\log n \to 0$ quand $n \to +\infty$.

THÉORÈME 4. – *Soit $(X_n)_{n\in\mathbb{Z}}$ un processus de Gegenbauer à k-facteurs Gaussien (5) d'écart-type σ, tel que $\max(d_1, \cdots, d_k) > 0$. Alors, F appartient au domaine d'attraction de Gumbel, i.e. :*

$$\mathbb{P}\{c_n^{-1}(M_n - d_n) \leq x\} \to \Lambda(x)$$

avec $c_n = \sigma(2\log n)^{-1/2}$ et $d_n = \sigma((2\log n)^{1/2} - \frac{\log\log n + \log 4\pi}{2(2\log n)^{1/2}})$.

1. Introduction and background

Extreme value theory is an area of statistics devoted to the development of models and techniques for estimating the behavior of unusual or rare events which are important in several domains as finance, internet traffic, meteorology, physic and biology. In this Note, we are interested to investigate this theory for a family of chaotic maps (the generalized tent family) representing deterministic dynamical systems with or without measurement noise, and for a particular class of long memory processes.

Throughout this Note, we denote $(M_n)_{n\in\mathbb{N}}$ the sequence of maxima defined by $M_n = \max(Z_1, \cdots, Z_n)$, $n \geq 1$, associated with a process $(Z_n)_{n\in\mathbb{N}}$ with marginal distribution function (d.f.) F, x_F the right end point of a d.f. F, and $g \in \mathcal{R}_\delta^{+\infty}$ a positive Lebesgue measurable function on $]0, +\infty[$ which is regular varying at $+\infty$ with index $\delta \in \mathbb{R}$.

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) random variables. The classical extreme value theory deals with the distribution of M_n as $n \to +\infty$, and its central result is the following theorem, derived by Fisher and Tippett [3] and first proved rigorously by Gnedenko [5], which gives the possible limiting distribution for maxima.

THEOREM 1. – *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables. If there exists some constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that:*

$$\mathbb{P}\{c_n^{-1}(M_n - d_n) \leq x\} \to H(x) \tag{1}$$

for some nondegenerate d.f. H, then H belongs to the type of one of the following three d.f.s:

Fréchet : $\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0,\ \alpha > 0 \end{cases}$

Weibull : $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x \leq 0,\ \alpha > 0 \\ 1, & x > 0 \end{cases}$

Gumbel : $\Lambda(x) = \exp(-e^{-x}),\ x \in \mathbb{R}.$

We say that the d.f. F of a sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ belongs to the domain of attraction of H, and we write $F \in \mathcal{D}(H)$, if (1) holds for some sequences $c_n > 0$ and $d_n \in \mathbb{R}$. Necessary and sufficient conditions are known for each type of limit, involving the behavior of the tail $1 - F(x)$ (denoted by $\overline{F}(x)$) as x increases.

Before defining the models on which we work, we recall the definition of extremal index which turns out to be useful in the sequel. Following Leadbetter *et al.* [11], a stationary process $(Z_n)_{n\in\mathbb{N}}$ has extremal index θ ($0 \leq \theta \leq 1$) if for every $\tau > 0$ there exits a sequence $(u_n)_{n\in\mathbb{N}}$ such that:

$$\lim_{n\to+\infty} n(1 - F(u_n)) = \tau \quad \text{and} \quad \lim_{n\to+\infty} \mathbb{P}(M_n \leq u_n) = e^{-\theta\tau}. \tag{2}$$

First, we consider the processes generated by a parametric family of chaotic maps defined by:

$$X_{n+1} = \varphi_\nu(X_n) = \begin{cases} 1 - (1 - 2X_n)^\nu, & 0 \leq X_n < 1/2 \\ 1 - (2X_n - 1)^\nu, & 1/2 \leq X_n \leq 1 \end{cases} \tag{3}$$

where $1/2 \leq \nu \leq 2$. This family contains the tent map and the logistic map, respectively for $\nu = 1$ and $\nu = 2$, and it is referred to the generalized tent family (see Hall and Wolff [9] and Lawrance and Spencer [10]). We also consider a noisy version of the system (3) for $\nu = 2$, defined by:

$$\begin{cases} Y_n = X_n + \varepsilon_n, & n \in \mathbb{N} \\ X_n = 4X_{n-1}(1 - X_{n-1}), & n \in \mathbb{N}^* \end{cases} \tag{4}$$

where $(\varepsilon_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables independent of $(X_n)_{n\in\mathbb{N}}$ with $\beta(a,b)$ density $(a, b > 0)$. The last model which we consider is the long memory Gaussian k-factor Gegenbauer process $(X_n)_{n\in\mathbb{Z}}$ defined by:

$$\prod_{i=1}^{k}(I - 2\upsilon_i B + B^2)^{d_i} X_n = \varepsilon_n \tag{5}$$

with $|d_i| < 1/2$ if $|\upsilon_i| < 1$, $|d_i| < 1/4$ if $|\upsilon_i| = 1$, $d_i \neq 0$, for $i = 1, \cdots, k$, and $(\varepsilon_n)_{n\in\mathbb{Z}}$ a Gaussian white noise (see Gray *et al.* [6] and Giraitis and Leipus [4]).

2. Maxima's limiting distribution of processes (3), (4) and (5)

In the Theorem 2 and 3, we provide the maximal domain of attraction of the marginal d.f.s of the processes (3) and (4), under the assumption that extremal index θ defined in (2) is different from zero.

THEOREM 2. – *Let $(X_n)_{n\in\mathbb{N}}$ be a process associated with the system (3) for a fixed ν, with invariant density f assumed to be continuous and strictly positive at $1/2$. If $(X_n)_{n\in\mathbb{N}}$ has extremal index $0<\theta\le 1$, then F belongs to the maximal domain of attraction of the Weibull distribution with index $1/\nu$, i.e.:*

$$\mathbb{P}\{c_n^{-1}(M_n-d_n)\le x\}\to \Psi_{1/\nu}(x)$$

with $c_n=(f(1/2)\theta n)^{-\nu}$ and $d_n=1$.

For instance, we get $F\in\mathcal{D}(\Psi_1)$, $c_n=(\theta n)^{-1}$ and $d_n=1$ for the tent process ($\nu=1$ in (3)), and we get $F\in\mathcal{D}(\Psi_{1/2})$, $c_n=(\pi/(2\theta n))^2$ and $d_n=1$ for the logistic process ($\nu=2$ in (3)).

THEOREM 3. – *Let $(Y_n)_{n\in\mathbb{N}}$ be a process associated with the system (4), assuming $(\varepsilon_n)_{n\in\mathbb{N}}$ has $\beta(a,b)$ density. If $(Y_n)_{n\in\mathbb{N}}$ has extremal index $0<\theta\le 1$, then F belongs to the maximal domain of attraction of the Weibull distribution with index $(2b+1)/2$, i.e.:*

$$\mathbb{P}\{c_n^{-1}(M_n-d_n)\le x\}\to \Psi_{(2b+1)/2}(x)$$

with $c_n=(\frac{4\Gamma(a+b)\theta n}{(2b+1)\pi\Gamma(a)\Gamma(b)})^{-2/(2b+1)}$ and $d_n=2$, where Γ denote the Gamma function.

Now, we provide a similar result for the process (5), remarking that its autocovariance function γ verifies the Berman's condition $\gamma(n)\log n\to 0$ as $n\to+\infty$ (see Berman [1]).

THEOREM 4. – *Let $(X_n)_{n\in\mathbb{Z}}$ be a Gaussian k-factor Gegenbauer process (5) with standard deviation σ, such that $\max(d_1,\cdots,d_k)>0$. Then, F belongs to the maximal domain of attraction of the Gumbel distribution, i.e.:*

$$\mathbb{P}\{c_n^{-1}(M_n-d_n)\le x\}\to \Lambda(x)$$

with $c_n=\sigma(2\log n)^{-1/2}$ and $d_n=\sigma((2\log n)^{1/2}-\frac{\log\log n+\log 4\pi}{2(2\log n)^{1/2}})$.

In establishing precisely the limiting distribution of their maxima, we have generalized the Theorem 1 of Fisher-Tippett [3] for the dependent processes (3), (4) and (5). Proofs are given in Section 4, and details can be found in Guégan and Ladoucette [8].

3. Estimation of extremal index

In Theorems 2 and 3 above, the extremal index θ defined in (2) appears in the normalizing constant c_n. Since we don't provide the value of θ analytically, we propose to estimate it. We choose the blocks method which involves dividing N given data into m blocks of length n and setting a high threshold u. A natural estimator of θ is then:

$$\hat{\theta}=n^{-1}(\log(1-K_u/m))/(\log(1-N_u/(mn)))$$

where K_u and N_u are respectively the number of blocks and the number of observations that exceed the threshold. We don't have the asymptotic normality of this estimator, but statistical properties are given in Smith and Weissman [12]. For the system (3) or (4), we simulate s realizations of the associated process in considering s different initial conditions X_0 in order to compute $\hat{\theta}$ and its standard deviation. Then, we obtain s estimated values for θ, and we compute $\hat{\theta}=(\sum_{i=1}^{s}\hat{\theta}_i)/s$, where the $\hat{\theta}_i$'s are the estimated values for a single realization, and $\hat{\sigma}_\theta=\sqrt{(\sum_{i=1}^{s}(\hat{\theta}_i-\hat{\theta})^2)/s}$. Thus, we can deduce $\hat{c}_n$, an estimate for c_n.

For instance, we consider the system (3) and fix $m=n=200$ and $s=100$. For $\nu=1$ and $\nu=2$, we respectively choose $u=0.9950$ and $u=0.9999$. We get $\hat{\theta}=0.98$, $\hat{\sigma_\theta}=0.02$

for $\nu = 1$, and $\hat{\theta} = 0.99$, $\hat{\sigma_\theta} = 0.01$ for $\nu = 2$. Similar estimations for the system (3), with different values of ν, and for the system (4), are provided in Guégan and Ladoucette [8]. Now, for specific densities f in Theorem 2, and specific values of a and b in Theorem 3, it is easy to compute $\hat{c}_n$.

4. Proofs

Proof of Theorem 2. – We suppose that $(X_n)_{n\in\mathbb{N}}$ has extremal index $0 < \theta \leq 1$. It is clear that $x_F = 1$. Following Hall and Wolff [9], we consider the pre-image of a general point $1 - x^{-1}$, $x > 1$, and we obtain:

$$f(1 - x^{-1}) = (2\nu)^{-1} x^{1-1/\nu} (f(1/2(1 + x^{-1/\nu})) + f(1/2(1 - x^{-1/\nu}))).$$

Then, we have the asymptotic $f(1 - x^{-1}) \sim \nu^{-1} f(1/2) x^{1-1/\nu}$ as $x \to +\infty$ i.e. $f(1 - x^{-1}) \in \mathcal{R}^{+\infty}_{1-1/\nu}$. Using a Karamata's theorem (see Theorem 1 page 281 in Feller [2]), we have $\overline{F}(1 - x^{-1}) \sim f(1/2) x^{-1/\nu}$ as $x \to +\infty$ i.e. $\overline{F}(1 - x^{-1}) \in \mathcal{R}^{+\infty}_{-1/\nu}$.

Thus, by Theorem 1.6.2. and Corollary 3.7.3. in Leadbetter *et al.* [11], we conclude that $F \in \mathcal{D}(\Psi_{1/\nu})$, $c_n = (f(1/2)\theta n)^{-\nu}$ and $d_n = 1$. The theorem is then proved.

Proof of Theorem 3. – Suppose that $(Y_n)_{n\in\mathbb{N}}$ has extremal index $0 < \theta \leq 1$. Let f, g and h denote respectively the invariant densities of $(Y_n)_{n\in\mathbb{N}}$, $(X_n)_{n\in\mathbb{N}}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$. It is obvious that $x_F = 2$. For $x \geq 1$, we have (see Guégan and Ladoucette [7]):

$$\begin{aligned} f(2 - x^{-1}) &= \int_{1-x^{-1}}^{1} g(y) h(2 - x^{-1} - y) dy \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_x^{+\infty} y^{-2} g(1 - y^{-1}) (x^{-1} - y^{-1})^{b-1} (1 + y^{-1} - x^{-1})^{a-1} dy. \end{aligned}$$

Then, using a Karamata's theorem, we obtain that $f(2-x^{-1}) \sim \frac{2\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{-b} g(1-x^{-1})$ as $x \to +\infty$. Since $f(2 - x^{-1}) \in \mathcal{R}^{+\infty}_{(1-2b)/2}$, we apply a second time a Karamata's theorem. We obtain that $\overline{F}(2-x^{-1}) \sim (2/(2b+1)) x^{-1} f(2-x^{-1}) \sim \frac{4\Gamma(a+b)}{(2b+1)\Gamma(a)\Gamma(b)} x^{-(b+1)} g(1-x^{-1})$ $\sim \frac{4\Gamma(a+b)}{(2b+1)\pi\Gamma(a)\Gamma(b)} x^{-(2b+1)/2}$ as $x \to +\infty$, i.e. $\overline{F}(2 - x^{-1}) \in \mathcal{R}^{+\infty}_{-(2b+1)/2}$.

Thus, by Theorem 1.6.2. and Corollary 3.7.3. in Leadbetter *et al.* [11], we conclude that $F \in \mathcal{D}(\Psi_{(2b+1)/2})$, $c_n = (\frac{4\Gamma(a+b)\theta n}{(2b+1)\pi\Gamma(a)\Gamma(b)})^{-2/(2b+1)}$ and $d_n = 2$, and the theorem is proved.

Proof of Theorem 4. – A stationary Gaussian process which verifies the Berman's condition $\gamma(n)\log n \to 0$ as $n \to +\infty$, belongs to the Gumbel domain with the same normalizing constants as in the i.i.d. case (see Theorem 4.3.3. in Leadbetter *et al.* [11]). Since it has a causal representation, the process (5) is Gaussian. From the expression of the asymptotic of its autocovariance function γ, given in Giraitis and Leipus [4], we obtain that:

$$\gamma(n) \leq \sum_{i=1,\cdots,k:d_i>0} A n^{2d_i^*-1}(B + o(1)) \quad \text{as} \quad n \to +\infty$$

with $d_i^* = d_i$ if $|v_i| < 1$ and $d_i^* = 2d_i$ if $|v_i| = 1$ $(i = 1, \cdots, k : d_i > 0)$, and where A and B are two finite positive constants. Since $2d_i^* - 1 < 0$, we have $n^{2d_i^*-1}\log n \to 0$ as $n \to +\infty$ $(i = 1, \cdots, k : d_i > 0)$, and then, we have $\gamma(n)\log n = o(1)$ as $n \to +\infty$.

Thus, the Berman's condition holds, and the proof is complete.

References

[1] Berman S.M., Limit theorems for the maximum term in stationary sequences, Ann. Math. Stat. 35 (1964) 502–516.

[2] Feller W., An Introduction to Probability Theory and Its Applications, Vol. 2, 2nd Ed., Wiley, New York, 1971.

[3] Fisher R.A., Tippett L.H.C., Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proc. Cambridge Phil. Soc. 24 (1928) 180–190.

[4] Giraitis L., Leipus R., A generalized fractionally differencing approach in long memory modeling, Lithu. Math. Journ. 35 (1995) 65–81.

[5] Gnedenko B.V., Sur la distribution limité du terme d'une série aléatoire, Ann. Math. 44 (1943) 423–453.

[6] Gray H.L., Zhang N.F., Woodward W.A., On generalized fractional processes, J.T.S.A. 10 (1989) 233–257.

[7] Guégan D., Ladoucette S., Invariant measures and second order properties for maps on $[0, 1]$, Preprint 01.07, University of Reims, France (2001).

[8] Guégan D., Ladoucette S., Extreme value theory of particular non linear processes, in preparation (2001).

[9] Hall P., Wolff R.C.L., Properties of distributions and correlation integrals for generalized versions of the logistic map, Stoch. Proc. Appli. 77 (1998) 123–137.

[10] Lawrance A.J., Spencer N.M., Statistical aspects of curved chaotic map models and their stochastic reversals, Scand. Journ. Stat. 25 (1998) 371–382.

[11] Leadbetter M.R., Lindgren G., Rootzen H., Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag, Berlin, 1983.

[12] Smith R.L., Weissman I, Estimating the extremal index, J.R.S.S. B 56 (1994) 515–528.